Ch6 : Statistiques descriptives - analyse des données

1. Caractéristiques de position : moyenne , médiane
2. Caractéristiques de dispersion : étendue, écart et intervalle inter-quartile
3. Utilisation de la calculatrice pour étudier une série statistique
4. Passage des effectifs aux fréquences et calcul des caractéristiques [1]
5. Calcul des Effectifs Cumulés Croissants ou Fréquences cumulées croissantes
6. Représentation d'une série statistique [2]

But :
Faire des études statistiques à partir des données de l'INSEE ou autre, et proposer une représentation pertinente.

1. à partir des fréquences ou des effectifs
2. distribuer la fiche d'aide "tableur" faite pour le collège

I Compléments de notions de collège indispensables sur les statistiques

a) **Vocabulaire** la **population** est l'ensemble sur lequel porte l'observation : on étudie un caractère bien précisé sur les **individus** de cette population (on collecte et on dépouille des données).

Un **échantillon** est une partie de la population.

b) **Type de caractère** lorsque le caractère étudié prend des valeurs numériques, le caractère est ..... , sinon le caractère est ..... .
- un caractère quantitatif est ..... lorsqu'il ne prend que quelques valeurs isolées ;
- un caractère est ..... lorsqu'il peut prendre toutes les valeurs d'un intervalle.

c) **Comparaison de populations ou d'échantillons de populations : intérêt réel des études statistiques**
On utilise des caractères de :
- **position** : moyenne, mode, médiane, quantiles (quartiles, déciles)
- **dispersion :comparaison de l'homogénéité des séries étudiées (graphiques, diagramme en boîte notamment)** .En comparant les valeurs suivantes : **étendue - intervalle - écart interquartile**

II Indicateurs de position

II-A Méthodes de calcul

II-A-a) Moyenne

noté $\bar{x}$, nombre qui,substitué à chaque valeur de la série, redonne la même somme.

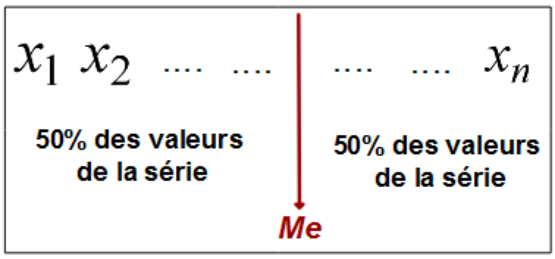
$$\bar{x} = \frac{\text{somme des } n_i x_i}{\text{somme des } n_i} = \text{somme des } f_i x_i$$

Valeur	x_1	x_2	...
Effectif	n_1	n_2	...
Fréquence	f_1	f_2	...

II-A-b) Médiane

Correspond à une valeur qui partage en **deux parties** (presque égales) la série statistique.

Définition : la médiane d'une série statistique est le nombre noté **Me** ou Q_2, tel que 50% au moins des individus ont une valeur du caractère inférieure ou égale à **Me**, et 50% au moins des individus ont une valeur supérieure ou égale à **Me**.



Calcul de la médiane :sa valeur est ...
- la valeur du milieu si **n est impair**
- la demi-somme des deux valeurs situées "au milieu" si **n est pair**

II-A-c) Mode :

II-B Étude 1 : savoir déterminer et interpréter moyenne et médiane

Ce tableau donne les salaires mensuels, en euro, des 185 employés d'une entreprise.

1. Déterminer, pour cette entreprise, le salaire moyen et le salaire médian.
2. Si les salaires étaient égaux, quelle somme recevrait chaque salarié ?
3. Peut-on affirmer ici que la moitié des salariés gagnent moins que la moyenne ?

Salaires(€)$\ldots_i$	1405	1480	1554	1870	2739	4215
Effectifs $\ldots_i$	40	15	81	35	9	5
E.C.C	...	...	...	...	...	...

E.C.C :.....

III Indicateurs de dispersion

III-A Méthodes de calcul

préalable : le premier quartile (Q_1) est la $\frac{N}{4}$ valeur (rang arrondi par excés) ; le troisième quartile (Q_3) est la $\frac{3N}{4}$ valeur (rang arrondi par excés).

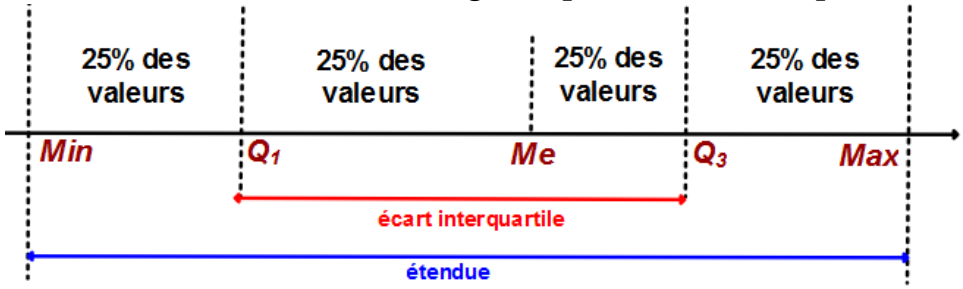
III-A-a) Étendue

Différence entre la plus grande et la plus petite des valeurs de la série (**maxX – minX**)

III-A-b) Ecart et Intervalle interquartiles

Ecart interquartile : différence $\mathbf{Q_3 - Q_1}$ **Intervalle interquartile :** $[\mathbf{Q_3 - Q_1}]$

Plus cet écart ou intervalle est grand, plus la série est dispersée.



III-A-c) Étude 2 : comparer deux villes Mexico et Barcelone

Voici les températures moyennes mensuelles de deux villes, en degrés celsius.

1. Pour chacune de ces deux villes :
   (a) calculer l'étendue de la série des températures
   (b) estimer la température moyenne annuelle
   (c) déterminer la médiane de la série
   (d) déterminer les quartiles Q_1 et Q_3 de la série

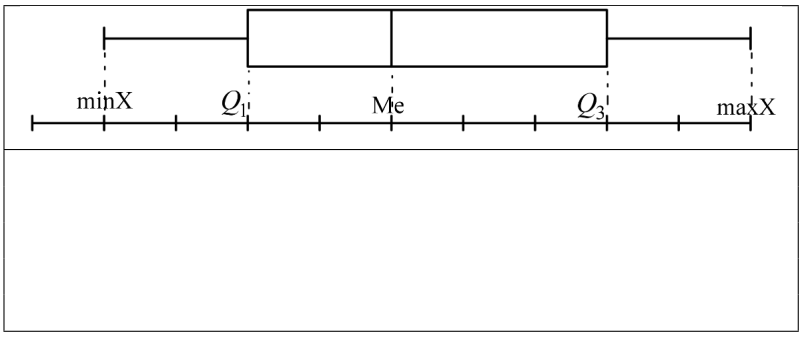
Mexico				Barcelone			
J	F	M	A	J	F	M	A
12,4	14,1	16,2	17,4	9,5	10,3	12,4	14,6
M	J	J	A	M	J	J	A
18,4	17,7	16,7	16,8	17,7	21,5	24,3	24,3
S	O	N	D	S	O	N	D
16,3	15,1	13,9	12	21,8	17,6	13,5	10,3

2. Quels calculs permettent d'affirmer :
   (a) " il fait plus chaud à Barcelone qu'à Mexico" ?
   (b) " dans ces 2 villes , la température est supérieure à 16°C pendant au moins la moitié de l'année " ?
   (c) "les écarts de températures sont moindres à Mexico" ?
   (d) "à Mexico, la moitié de l'année, il fait environ entre 14°C et 17°C" ?

III-B Étude 3 : diagramme en boîte : réalisation et utilisation

Reprendre les données de l'étude 1, et déterminer l'étendue et l'intervalle inter-quartile. Réaliser le diagramme en boîte en utilisant le modèle ci-contre. **Commenter.**

. . . . . .
. . . . . .
. . . . . .
. . . . . .



III-C Étude 4 : détermination des indicateurs à partir des fréquences

Reprendre les données de l'étude 1 :

1. calculer les fréquences arrondies au millième près, et les fréquences cumulées croissantes.
2. Quel est le pourcentage de salaires inférieurs ou égaux 1870 € ?
3. Calculer moyenne, médiane et quartiles de la série en utilisant les fréquences.

Salaires(€)$\dots_i$	1405	1480	1554	1870	2739	4215
Effectifs $\dots_i$	40	15	81	35	9	5
Fréquences	...	...	...	...	...	...
F.C.C	...	...	...	...	...	...

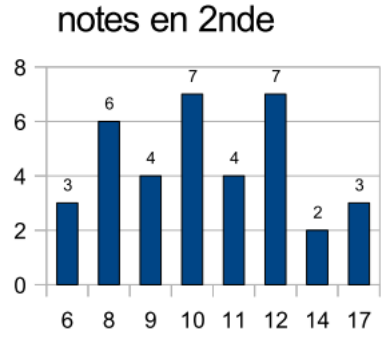
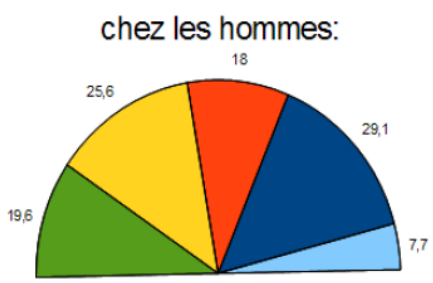
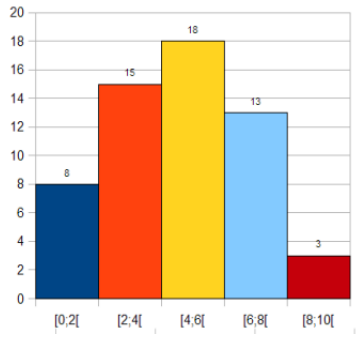
Méthode de calcul de la fréquence : .....

.....

.....

IV Représentations graphiques pertinentes

IV-A Selon le type de caractère ...

Caractère quantitatif discret ou qualitatif	**Tout type de caractères**	**Caractère quantitatif continu**
Diagramme en bâtons	**Diagramme semi-circulaire ou circulaire**	**Histogramme**
		

IV-B Faire une représentation pertinente des données de l'étude 1

IV-C Étude 5 :Courbe des F.C.C (fréquences cumulées croissantes)

Un distributeur automatique de café propose des expressos.

81	82	85	83	82	87	83	84	85	84
84	81	83	86	84	80	79	87	85	80
81	82	85	82	79	80	86	89	83	89

Une pesée portant sur 30 expressos a donné les masses suivantes (en grammes) de café utilisé (voir tableau ci-contre).

Masse (en g)	[79;82]	[82;85]	[85;88]	[88;91]
Fréquence				
F.C.C				

1. Compléter le deuxième tableau ci-contre par les valeurs approchées au centième des fréquences et par les fréquences cumulées décroissantes (F.C.C).
2. Représenter la courbe des FCC (graphique à faire en Annexe B).
3. Par lecture graphique, recopier et compléter : "75% des expressos contiennent plus de .....g de café. "

Annexe A : Mode d'emploi de la calculatrice

• La touche stats permet d'accéder au menu des statistiques.

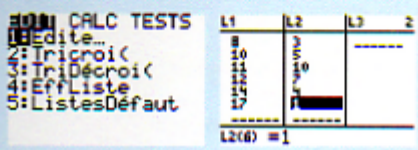
Entrer des données

• Sélectionner le menu « EDIT » puis « Edite... » et appuyer sur entrer.

• Saisir les données statistiques dans les listes.

Exemple

On saisit les notes dans la liste L_1 et les effectifs correspondants dans la liste L_2.



• Appuyer sur entrer.

• **Après avoir vérifié que *n* indique le bon nombre de données,** la médiane et les quartiles s'obtiennent en demandant le bas de l'affichage avec la flèche.

• Pour quitter la fenêtre des résultats, appuyer sur la touche annul.

Obtenir les paramètres statistiques

• Dans les menus de stats, sélectionner « CALC » puis «Stats 1-Var » et appuyer sur entrer.

• Saisir les noms de listes où se trouvent les valeurs x_i du caractère étudié et les effectifs n_i correspondants.

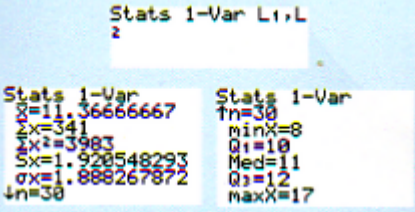
Exemple

Si les données ont été saisies dans les listes L_1 et L_2, appuyer sur : 2nde 1 , 2nde 2 .

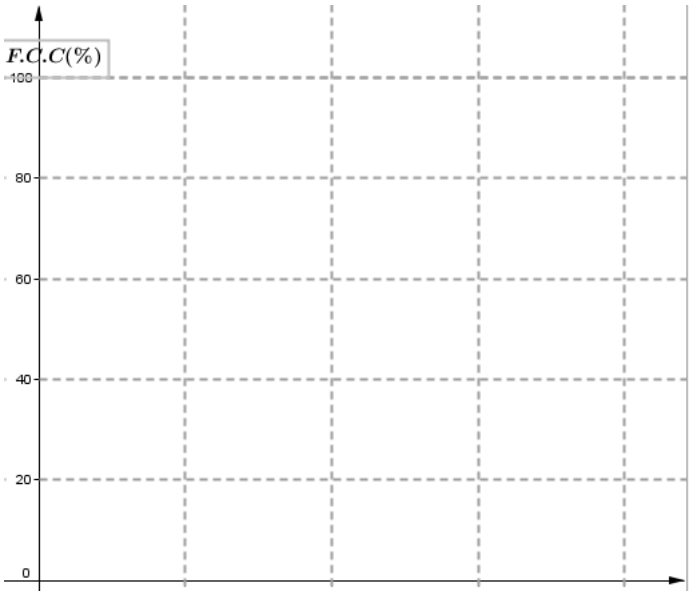
Remarque

Si les effectifs associés à chacune des valeurs du caractère sont tous 1, il suffira d'appeler la seule liste contenant les valeurs du caractère.

Exemple



Annexe B : Courbe des F.C.C de l'étude 5



Annexe C : Comparer deux séries :

Pour tester l'efficacité d'un médicament destiné à faire baisser la proportion d'une substance dans le sang, on compare les résultats d'analyse dans le groupe de patients A recevant le médicament et le groupe B recevant un placebo :

GA	1,9	1,6	1,3	1,5	1,5	1,8	1,5
	1,8	2,1	1,7	1,7	1,7	1,5	1,4
	1,7	1,6	1,6	1,9	1,6	1,5	1,7
	1,6	1,8	1,8	1,6	1,5	1,8	1,5
GB	1,9	2,1	2,0	2,0	1,8	2,1	2,0
	1,8	1,7	2,1	1,9	1,9	2,1	1,8
	1,7	1,9	1,9	1,8	1,9	1,7	2,0
	1,9	2,4	2,4	1,9	1,8	1,5	2,1

Suite Annexe C :

(a) Proposer un schéma permettant de comparer les deux séries

(b) Peut-on considérer que le médicament est efficace ?